DOCUMENT RESUME

ED 342 209 FL 019 252

AUTHOR Ingram, D. E.; Wylie, Elaine
TITLE Developing Proficiency Scales for Communicative Assessment.
PUB DATE 5 Dec 89
NOTE 21p.; Abridged version of a paper presented at the National Assessment Consultation for the National Assessment Framework for Languages at Senior Secondary Level (Sydney, Australia, December 5, 1989).
PUB TYPE Guides - Non-Classroom Use (055) -- Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Communicative Competence (Languages); Foreign Countries; *Language Proficiency; *Language Tests; *Rating Scales; Secondary Education; *Second Languages; *Test Construction; *Testing
IDENTIFIERS *Australia

ABSTRACT

A discussion of language testing focuses on how proficiency scales are constructed and applied to communicative language assessment in Australian secondary schools. A variety of rating scale types are described and distinguished according to how they are constructed, what they attempt to measure, and what criteria they contain. A more detailed analysis of testing criteria provides models of proficiency in writing and listening. The process of constructing a scale is examined, with examples from the Australian Second Language Proficiency Ratings and other commonly used proficiency scales. The ways in which learner proficiency can be rated are enumerated, including analytic matching, interpretation of indirect and semi-direct test scores, checklists, non-structured observation, structured interview, criteria subscales, and self-rating. Five key areas in which proficiency scales have a role in communicative assessment are examined, including assessment of language use in carrying out communicative tasks, syllabus design, measurement of individuals' language learning progress, formative and summative assessment, and tracking. Finally, identification of appropriate scales for secondary school communicative language courses is discussed, particularly as they relate to grading. The rating scale for speech from the Queensland Senior Italian syllabus is appended. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED342209

Developing Proficiency Scales for Communicative Assessment

D. E. Ingram, Professor of Applied Linguistics and
Director, Centre for Applied Linguistics and Languages
and
Elaine Wylie, Director of Studies
Institute of Applied Linguistics,
Kelvin Grove Campus, Queensland University of Technology

This is an abridged version of a paper of the same title presented at the National Assessment Consultation for the National Assessment Framework for Languages at Senior Secondary Level, (NAFLaSSL), Sydney, 5 December, 1989.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Ingram D.E
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

FL019253

BEST COPY AVAILABLE

PROFICIENCY:

The main purpose of this paper is to consider how proficiency scales are constructed and applied to communicative assessment in Secondary Schools. The concept of proficiency has been discussed at length elsewhere, especially in Ingram 1985. For the present purposes, it is sufficient to say that, especially in papers on the Australian Second Language Proficiency Ratings (ASLPR), the present writers have defined proficiency as entailing notions both of the sorts of tasks that can be carried out in the language and the ways in which those tasks are carried out, i.e., the sorts of language forms that occur. We also distinguish between language proficiency and communicative competence, at least in the popularist sense of "communicative competence", since the ability to communicate entails more than the ability to mobilize language forms and practical knowledge of how to carry out a language task [e.g., Ingram 1985:226-227]. It is also fundamental to the point of being trivial to distinguish between a learner's language proficiency and his or her ability to perform activities (such as academic studies, a vocational task, social interaction, seduction, or winning elections in Queensland) in which language may occur but is only one of many factors that determine the outcome of the event: for such purposes, one may, for example, assess whether the learner has enough language to carry out the task without language proficiency being an inhibiting factor but clearly it is evident to the point of being trivial to assert that such a favourable assessment does not mean that the person will succeed in the degree programme, seduce the target person, or win a State election.

SCALES:

Amongst the two columns of definitions the *Shorter Oxford English Dictionary* provides for "scale" is "an apparatus for weighing" and "a set or series of graduations...used for measuring distances etc". A proficiency scale selects certain graded criteria against which to "weigh" learners' ability to use the language, it selects criteria that can be graduated between two points so that learners' ability to use the language or some relevant aspect of their language can be matched against the criteria themselves. Generally a number of intermediate points are selected and criteria assigned to them so that a "set or series" of graduated steps is provided between the two end points. It is generally assumed that the gradation in the criteria and in the intermediate points to which they are assigned reflects progression in mastery of the language or some aspect of it so that the match indicates the learners' stages of development. In the case of proficiency scales such as the ASLPR, the scale consists of graded descriptions of language behaviour so that, in arriving at a match between learners' language behaviour and the scale, one is saying what the learners' language behaviour is like, hence the sorts of tasks they can carry out in the language and how they carry them out.

A variety of different types of proficiency or proficiency-related scales may be distinguished according to how they are constructed, what they attempt to measure, and what criteria they contain. The following can serve as examples:

Whole vs. Part: Proficiency scales may relate to the whole span of proficiency development from zero to native-like or the scale may relate to only a part of that development. In the case of the ASLPR and the International English Language Testing System (IELTS), the span envisaged is from zero to native-like. Though it is difficult to arrive at an indisputable definition of "native-like", here it is assumed to be that point where the

learner's language behaviour is indistinguishable from that of a similarly educated first-language speaker of the same sociocultural dialect.

Serial vs. Threshold: A proficiency scale may seek to provide a series of intermediate points between two levels (e.g. zero and native-like) or it may select a threshold level which is described with possibly more cursory (if any) descriptions of behaviour above and below this threshold. The threshold level is set in relation to the use to be made of the scale, e.g., if the scale is to be used to decide whether a learner has enough proficiency to practise a certain vocation or to undertake a course of study, the threshold description will select criteria to describe the minimum proficiency required for those purposes and the testing procedure has only to determine whether the learner's proficiency and the focused criteria fall at, below or above the threshold: this is generally a simpler assessment procedure than providing a comprehensive rating of a learner's proficiency.

General vs. Special Purpose: A proficiency scale may seek to provide an assessment of learners' general proficiency [see Ingram 1985: 223-224] or their ability to use the language in some special purpose area of the language (e.g., in English for Civil Engineering, English for Nurses, etc). The ASLPR aims to assess general proficiency but special purpose versions of the ASLPR are also in preparation for use in assessing special purpose proficiency in designated areas. The IELTS provides measures of general proficiency in Listening and Speaking and, for some candidates, in Reading and Writing and special purpose proficiency in three very broad areas of academic Reading and Writing (i.e., in English for academic purposes in the Life and Medical Sciences, in the Arts and Social Sciences, and in the Physical Sciences and Technology).

Task-only vs. Total behaviour: A scale may select as criteria the tasks that learners can carry out at graduated levels or it may attempt to select criteria that provide a more comprehensive picture of the learners' total language behaviour. The ASLPR, FSI Scale and ACTFL attempt to do the latter, as do the IELTS Band Scales but using much less comprehensive descriptions. The problem with task-only scales such as some purist early functional courses attempted is that a task may be carried out using a variety of language forms or none at all and so a task-only assessment is not very informative so far as the learner's use of the language is concerned.

General Proficiency vs. Graded Objectives: A scale may attempt to provide descriptions of the learners' general proficiency at each of the graduated levels or it may adopt the sort of graded objectives approach popular in the late 1970s when a graduated series of tasks were selected and their performance described using criteria considered relevant for consideration in the light of the course being taught. The graded objectives may be relevant to general proficiency but they need not be, especially if, as sometimes occurred in graded objectives courses, the course neglected the systematic development of the language and its creative use in favour of task-achievement by, for instance, rote-memorization.

Macroskill-specific vs. Overall: Scales may seek to differentiate the four macroskills, as does the ASLPR, in the belief that different macroskills may differ in their rate and direction of development and in the criteria that must be observed while other scales (e.g., the IELTS Overall Band Scale) seek to describe "general language development" possibly in the sense in which Oller used the term "general" in the 1970s or, in any case, to refer to some generalized language development irrespective of the features pertinent to any specific macroskill [see Ingram 1985: 224-225 for a brief discussion of

unitary vs. macroskill specific scales and of Oller's views; see Oller 1976, amongst many others, for a discussion of his notion of "general" proficiency].

Absolute vs. Global: Some scales (e.g., the FSI Scale) describe themselves as "absolute" in the sense that all criteria within a descriptor must be observable before learners are assigned that particular level as their proficiency rating. In contrast, the ASLPR claims to be a global scale in the sense that the descriptors provide a global description of the learners' behaviour at any particular level. The introductory paper to the ASLPR states:

> "It must be stressed that it is the observer's global assessment that is important and the definitions provided are not to be regarded as checklists. In this respect, the ASLPR differs from the FSI Scale. In the ASLPR, the definitions provide an overall picture of language behaviour at each level, and the learner is assigned to the one which his or her performance most closely resembles. This approach recognizes that language is highly complex, that it may develop at slightly different rates in different directions [e.g., syntax as against phonology, phonology as against discourse structuring, etc] and that, therefore, minor but compensating variations may occur within the total picture of the learner's behaviour. In some cases maverick development of a component skill (especially one that is strongly influenced culturally) may occur without significantly influencing the level at which the learner is assessed. [Ingram 1979: 9-10]

The global approach would seem to make assessment more subjective but it is necessitated by the inherent complexity of language behaviour and any loss of reliability is compensated for by other features of the rating procedure and, in particular, by the better account that it takes of the complex nature of language itself [see Ingram 1979: 10; 1985: 222-223].

Analytic vs. Holistic: To some extent this contrast is similar to that between analytic and synthetic tests [see Ingram 1985: 229] though the distinction is not often clear-cut in proficiency scales. Some set out to include quite specific statements about features (e.g., grammatical forms) that occur at each level described while others seek to make more generalized statements that, for instance in the case of syntactic development, describe the overall appearance of sentences rather than specific syntactic items that initially emerge at a particular level. The ASLPR aims to provide generalized statements but to assist in their interpretation by giving examples of the forms as well as the tasks that might be observed at a particular level. The problem with proficiency scales that make quite explicit statements about features of, for example, syntax, phonology or functions that occur at each level is that the complexity of language may mean that a rating assigned using such criteria may be of little relevance to the learner's practical proficiency because of the complexity of language and the fact that more or less development in one feature may be compensated for by less or more development in other features, i.e., as a measure of syntactic development it may be accurate and useful for many purposes (not least for diagnostic purposes and for streaming into formal classes) but it may be of little use as a statement of practical proficiency.

Fully Empirical vs. Anticipated 'Washback': Most scales simply describe language as it is observed in a particular group of learners and then distilled. There will be some principles of organization which aim to help the person making the judgement. For example, the criteria included may be restricted in number, or some may be highlighted by generous use of exemplars.

There may, however, be a case for deviating from this purely empirical approach, for describing what is deemed to be desirable as opposed to what is actually observed. In scales devised by Wylie for the Queensland Senior Italian syllabus, (see Appendix 1 for Speaking Scale) there was a conscious decision to depress slightly the standard generally observed for certain elements. For example, the standards given in the 'Appropriateness' paragraphs in the level descriptions for Very High Achievement and High Achievement in speaking are slightly lower than those generally observed in Very High and High Achievement students at the end of Year 12. Experienced teachers consulted had agreed that a relaxation of the standards, particularly where register rather than communication of meaning is the issue, could result in less inhibition and more spontaneous, fluent speech.

CRITERIA:

The contrasts that have just been made between different types of scales relate in most instances to the criteria that are chosen for inclusion in the descriptors.

Criteria-based assessment in Queensland requires that the identification of criteria be very 'up-front' in the syllabus and work programs, much more so than in, for example, the ASLPR. This brings with it the need for a more desciplined approach to the organization of the scale.

Sadler, in one of the Queensland Board of Secondary School Studies Assessment Unit papers, suggested a schema for organizing a potentially very large number of criteria into a hierarchy whereby very specific criteria which may or may not be relevant to all tasks can be subsumed under a workable number of more general criteria [Sadler, 1986]. Wylie applied this principle to LOTE syllabuses, with the aim of establishing a working level in the hierarchy for each macroskill which would be both comprehensive and universally relevant. The resulting schemata are slightly more complex than Sadler's.

The Writing schema, shown in Figure 1 below, has an initial dual focus at the working level on the traditional areas of content and form, the 'what' and the 'how'. The separate identification of these should not be taken to suggest that there by any compartmentalization of approach in assessment activities. This would, in fact, be futile to attempt. We cannot express ideas except through some symbolic forms such as language, and language forms unrelated to ideas are not communication. But we can focus on, and make separate judgements about each, just as we can about the distinct skills of Speaking and Listening in a task such as a conversation, which involves them both.

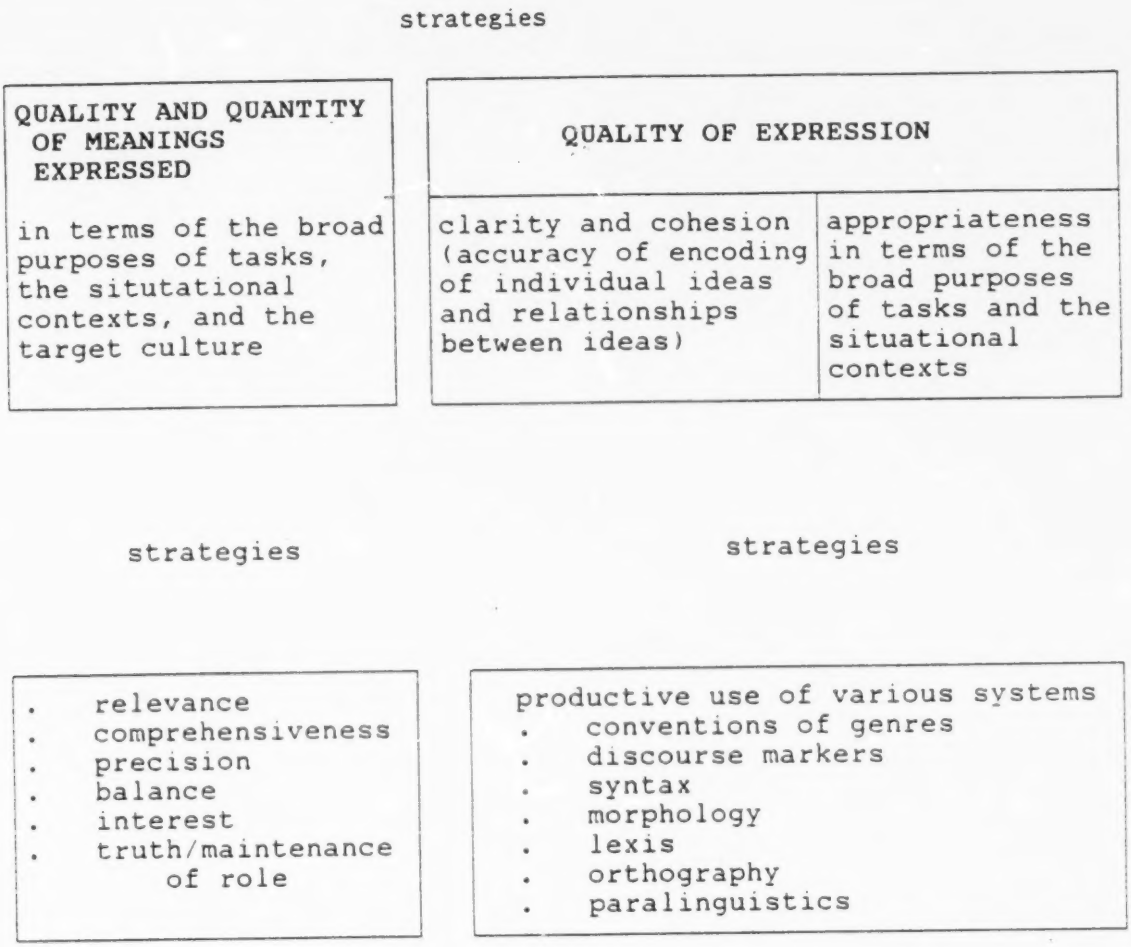
strategies
QUALITY AND QUANTITY OF MEANINGS EXPRESSED
in terms of the broad purposes of tasks, the situtational contexts, and the target culture
QUALITY OF EXPRESSION
clarity and cohesion (accuracy of encoding of individual ideas and relationships between ideas)
appropriateness in terms of the broad purposes of tasks and the situational contexts
strategies
strategies
. relevance
. comprehensiveness
. precision
. balance
. interest
. truth/maintenance of role
productive use of various systems
. conventions of genres
. discourse markers
. syntax
. morphology
. lexis
. orthography
. paralinguistics

Figure 1

'Coherence' refers to "connectivity of underlying content", the way the ideas and the relationships between ideas in the 'text world' match those of the real world of the listener or reader, and 'cohesion' to "connectivity of the surface" (see de Beaugrande and Dressler, 1981).

Strategies are not elaborated. This is not meant to suggest that they are unimportant, but this model is concerned more with product than process, and successful strategies are often not observable.

The model for Speaking is very much the same as that for Writing, with 'fluency' added under QUALITY OF EXPRESSION as a further criterion at the working level.

The tripartite division into different types of information in the Listening schema shown in Figure 2 below and in Reading has links with traditional models of language proposed by Jakobson and Buhler (see Jakobson 1960, pp 352-7).

PROFICIENCY IN LISTENING

strategies

QUALITY OF SENSE MADE OF TEXTS (COHERENCE RECREATED)		
accuracy and fluency of decoding of individual ideas and relationships between ideas	validity of perception of author or speaker's purposes	validity of perception of affective meanings (particularly attitudes)

strategies

receptive use of various systems
. conventions of genres . discourse markers . syntax . morphology . lexis . phonology . paralinguistics

Figure 2

To obtain an acceptable degree of comparability of achievement levels between schools, the performance-related criteria discussed above need to be complemented by another set of criteria, those related to the difficulty of the task performed. These can be classified into two types - those which relate to the inherent difficulty of the task, and those which relate to assistance provided for the student in performing the task. For example, with writing tasks, Wylie has suggested the following:-

INHERENT DIFFICULTY OF TASK	ASSISTANCE PROVIDED WITH TASK
. comlexity and precision of individual meanings . number and types of discourse links . amount of extra-linguistic processing required . familiarity of student with purpose(s), genre, context, meanings (for idiosyncratic or cultural reasons)	. amount of rehearsal by the student of the actual task, similar tasks, or parts of the task . provision of dictionaries and other reference material . provision of contextual material (e.g. cue(s) in the target language)

CONSTRUCTING PROFICIENCY SCALES:

The type of scale chosen and the nature of the criteria to be included in it determine to some extent how one will go about constructing a scale. If the scale is to be highly analytic and itemize, for example, syntactic forms, then one can identify either from developmental psycholinguistics or by testing exactly what syntactic forms appear in the behaviour of learners at the designated levels.

In the case of the ASLPR, the IELTS Scales, and the ACTFL Scale, a different approach was adopted even though all would claim to have sought to take account of the literature of developmental psycholinguistics at least to the extent of seeking to ensure that the broad descriptive statements were compatible with what was known from the more analytic item-related studies. These scales (the ASLPR and the others) seek to capture the intuitions that native-speakers have about their language and its development, make them more concrete and assessable using descriptions of the language features that are observable, and trial them and flesh them out by actual observations of learners.

As described in Ingram 1979 [p. 7], the starting-point for the ASLPR was the FSI Scale and, for the ACTFL Scale, the ASLPR and others. The FSI Scale was chosen for a number of reasons (spelt out in Ingram 1979: 7) but the principal aim was to draw on the intuitions of experienced teachers of a variety of languages in order to describe language behaviour at a number of proficiency levels between zero and native-like. Drawing on the FSI Scale as a starting point enabled a great deal of experience and experiment in describing proficiency and its development to be added to that of the researchers working on the ASLPR. As this early and more or less intuition-based scale was written, it was subjected to extensive trialling, i.e., the researchers interviewed and assessed large numbers of learners at all proficiency levels from zero to native-like in order to assess the consistency and coherence of the descriptors and to elaborate them. In other words, the earlier partly intuitive descriptors were tried out with real-life learners and the descriptors themselves were assessed and elaborated empirically. Formal trials were held in Australia, China and the United States to assess the validity and reliability of the scale when used with adult and adolescent learners of ESL and of three other languages. This formal trialling

essentially assumed that, if the series of descriptors making up the scale described real language development, if the descriptors were coherent (i.e., described features of the language that actually did generally co-occur), and if the descriptors were comprehensible and manageable, then teachers trained to use them could interpret them consistently, apply them reliably to rating the same learners, and agree on the ratings assigned.

The developmental process described has, in fact, continued over the years as other research projects were undertaken and as the scale was used in practice. A similar approach has been commenced with the IELTS Band Scales though, to date, they have not gone beyond the intuitive stage of drawing on experience, teachers' insights, and other scales in order to arrive at descriptors. It is vital that, in the immediate future, these band scales be tried out with large numbers of learners, the descriptors re-assessed and re-written in the course of the face-to-face interviews, and that they be formally trialled.

RATING OR ASSIGNING LEVELS TO LEARNERS:

There are a number of different ways in which learners may have their proficiency rated.

Analytic Matching: In the case of more analytic scales or where a researcher believes that specific features in the language (e.g., syntactic forms) can be correlated with particular proficiency levels, it may be possible to use indirect and semi-direct tests types to objectively relate learners to the scale. However, the problem of the complexity of language development referred to earlier remains; for example, unless one observes the learner's actual language behaviour, a rating based solely on knowledge of a particular syntactic item may not accurately reflect the learner's practical proficiency.

Interpreting Indirect and Semi-direct Test Scores: The IELTS uses direct tests of Speaking and Writing (i.e., tests in which language behaviour is observed and rated against a band scale) and indirect and semi-direct test types in Listening and Reading. In the latter cases, scores have to be re-interpreted in terms of band scales in ways that are not yet, at least in the view of the present writers, convincingly resolved but probably by some psychometric process yet to be determined. In other words, ratings may be assigned by correlating scores on indirect and semi-direct tests with the scale levels.

Checklist: A checklist of criteria and related standards may be used so that if some or all of the criteria for a level are marked off (when observed in a formal interview or in real life), that is the level assigned. This is intended to be the approach with the FSI Scale though it appears that, for the reasons indicated earlier in implied criticism of the checklist approach, it is not always used in practice [John Clark, personal communication, 1981].

Non-structured Observation: Learners may be observed over a long period of time either in class or in real-life activity and their proficiencies rated against a scale. Though this would seem to provide a highly practical measure of the learners' proficiency, it may be both invalid and unreliable since, for example, the rater may not have observed the learners in situations that extend their ability to use the language and so the rating may not be a true measure of their proficiency. In addition, where such ratings are based on classroom observations, the teacher's ratings may be distorted by other

factors such as the learners' class behaviour, their formal knowledge of the language which is often the focus of class activity, and their earlier stages of development.

Structured interview: To overcome the last problem, scales such as the ASLPR use a structured interview to elicit the learner's maximum language behaviour which is then rated against the scale. In the case of the ASLPR, the interview is described in broad stages in order to allow the interviewer flexibility to vary the interview for certain purposes so long as the learner's maximum language behaviour is elicited. In the case of the IELTS Speaking test, the interview is more controlled in order to try to increase reliability in situations where relatively untrained interviewers in a great variety of situations are conducting the interviews though the same purpose applies in this as in the freer interview type, i.e., to elicit the learner's maximum language behaviour in order to match it against one of the scale descriptors.

Criteria sub-scales: In assessing scripts in the IELTS Writing test, examiners use sub-scales to rate selected criteria which are differentially weighted and then reduced to a rating on a single band scale. The problem in such an approach is to resolve the weightings assigned to the sub-scales and the best way to translate the sub-scale scores into the single band scale score. It has also yet to be shown that this approach is any more valid and reliable than the use of the single band scale, which would seem to be simpler, less time consuming, and no more prone to (or at least with fewer opportunities for) subjectivity or impressionism.

Self-rating: Self-assessment is a variation of the non-structured observation discussed above. Wylie has developed several simplified versions of the ASLPR for different types of learners to assess their own proficiency, and limited trailling has suggested a high degree of reliability. The latest version. of which the scale for Speaking is given in Appendix 2. was devised for Queensloand teachers who are being withdrawn from the classroom for 15 week intensive Japanese courses.

PROFICIENCY SCALES IN COMMUNICATIVE ASSESSMENT:

Proficiency scales like the ASLPR and ACTFL Scale designed to measure general proficiency as it develops from zero to native-like can serve a variety of purposes in communicative language courses and in communicative assessment. The present authors have written extensively on the place of proficiency scales and proficiency assessment in language courses [e.g., Ingram 1985] and here only some key areas will be selected for discussion.

First, it is necessary to recall the distinction made earlier between proficiency and communicative competence. If one sees communicative methods leading to the development of communicative competence in the loose sense of the ability to communicate, then a proficiency scale does not measure this ability for the simple reason that other factors intervene (including personality, knowledge of the field to be discussed, intelligence and education, the ability to systematize knowledge and see and express logical relationships, the learner's relationship with the other participant in the interaction and the latter's willingness to accommodate less than native-like language forms). Indeed, granted the number and variety of factors other than language that determine the ability to communicate (including factors related to the other interlocutor rather than to the learner, it is questionable whether language teachers should attempt to test or develop tests to assess communicative competence. If, however, a teacher were to aspire to test

communicative competence do so, then different scales would be required from proficiency scales, scales that seek to include the additional factors that compose communicative competence. In brief, it is entirely appropriate that language courses seek to develop the ability to communicate but it is probable that adequate tests of this ability cannot be developed and, because, to be complete, they would have to make statements about an interlocutor whose characteristics are irrelevant to the learner, they are not appropriate to language testing; rather, teachers should seek to assess a learner's ability to use the language in carrying out communication tasks, i.e., that they should seek to measure proficiency.

Second, because proficiency scales trace the development of learners' language from zero to native-like, they can make significant contributions to syllabus design. Scales like the ASLPR focus attention on practical language skills thus encouraging teachers to re-think what they are about in teaching the language. For instance, because proficiency scales focus on practical ability to use the language rather than on formal knowledge about the language, the introduction of proficiency scales may have a highly beneficial washback effect on teaching. This was, in fact, a major consideration in designing the IELTS. In addition, proficiency scales provide a firm basis on which to set long-term goals and objectives, they allow the learners' entry levels to be clearly specified, and they show the probable path of development of any learner, e.g., if a learner is at 1+, then the ASLPR shows that a major pattern of change in the language over the next period will be the complexification of the language, stimulated as learners try to express themselves in the course of social interaction. It would seem probable that, if courses were designed to take into account the natural developmental trends (especially the major parameters of change identified in the scale), then learning would be made most effective.

Nevertheless, it has to be stressed that a proficiency scale is not itself a syllabus: it provides a developmental description of proficiency as it changes from zero to native-like, it identifies a series of key parameters of change as proficiency develops, and it contains a series of static descriptions of language behaviour at various points between zero and native-like but it is not itself a syllabus: there is a logical and actual difference between a proficiency level and how a learner reaches that proficiency level. Thus, for instance, learners at ASLPR Level 1 could not attain Level 2 proficiency by simply memorizing examples of language behaviour at Level 2. In brief, a proficiency scale is not a syllabus but, in that it provides a means of specifying entry levels and goals and objectives and in that it reflects paths of development, it provides an overarching framework within which syllabuses and courses can be designed. In addition, its clear specification of proficiency levels allows series of courses to be made more coherent, it facilitates the movement of learners between courses, and it provides a means of assessing the suitability of activities and materials to the learners' stage of development.

The approach to syllabus design for ELICOS courses at the Institute of Applied Linguistics at the Brisbane College of Education exemplifies this. The genre or text-type is the main unit of organisation for Level 1 and above. At each level of the ASLPR there are references to particular genres which the learner can use to some degree. For learners who are (say) at Level 1, it is appropriate to focus on genres which are identified or implied at 1+, since 1+ language behaviour is the goal for these learners. The table below shows some of the genres which are suggested for general English courses for various levels.

Genre / Level	1	1+	2	2+	3
anecdote	very simple or simplified	simple			
display ad					
editorial			simplified on familiar topic	very simple on familiar topic	simple on familiar topic
feature article in magazine or newspaper	very simple or simplified	simple			
interact-ional conver-sation	very simple f to f	simple f to f very simple on phone	f to f simple on phone		
news item on radio	very simple or simplified	simple	moderately simple	including actuality	including actuality
transac-tional conver-sation	simple f to f, very simple on phone	f to f very on phone		on phone	on phone

Table 1

Third, one of the important potential uses of proficiency scales in the context of syllabus writing and course planning is to enable more adequate information on learners' rates of progress through the language to be accumulated. At present, little firm information is available on how rapidly different learners with different educational and other backgrounds and in different types of courses move from one proficiency level to another. If such information were available, it would enable courses to be planned on a more informed basis and would reduce the unrealistic expectations that many people (not least finance-managers, politicians, other administrators, and often teachers) have of what proficiency changes can be achieved in any particular course, e.g., it might lead the Federal Government to re-consider its unwillingness to fund four-year degree programmes while, at the same time, urging that tertiary students be encouraged to develop proficiency in a language while undertaking other degree courses.

Fourth, a proficiency scale may be used in formative and summative assessment. In formative assessment, it may be useful in judging the suitability of other test items, test types and texts to the learners' stage of development. For instance, the specifications for the IELTS test use the IELTS band scales to specify the levels of texts and activities that can be used in the various sub-tests. A proficiency scale may also be useful in formative assessment to assess whether the learner is developing the practical and functional skills that a course may aim at. Nevertheless, generally in formative assessment, the teacher is interested in finding out whether the elements of the language he or she is teaching are being internalized by the learner and, for this purpose, more analytic indirect and semi-direct tests will generally be more useful.

For summative purposes in a course aimed at developing practical language skills, a proficiency scale is very useful in that it provides a measure of the learners' practical proficiency. However, it has to be stressed that proficiency changes slowly and there is little point in using a proficiency scale and the associated test procedures at frequent intervals and, in general, they could not be advocated at more frequent intervals than about 250 hours, i.e., at the beginning and main exit points in the usual Secondary School foreign language class or at the start and finish of a full-time ten-week ESL module for overseas students.

In this context, it is useful to distinguish course attainment from general proficiency. Language teaching necessarily assumes that learners will be able to generalize whatever they learn in class to other situations and other purposes but often in school assessment the teacher is interested in the extent to which the learner is learning what the teacher is teaching. In other words, he or she is interested in the learner's attainment on the particular course being taught. For real-life purposes, however, one is more likely to be interested in the extent to which the learner can use the language for practical purposes and so one would wish to measure the learner's general proficiency using a proficiency rating scale. Ideally, therefore, at the end of a course (especially if one wishes to evaluate the success of the course itself), one would provide both a course-related achievement test using items relevant to what was taught and what one wanted to find out about and a test of general proficiency using a proficiency rating scale.

Wylie's proposed scales for Senior Italian in Queensland (see Appendix) are first and foremost achievement scales but, in keeping with the aims of the course, there is a strong element of general proficiency, particularly at Very High and High Achievement Levels, with references to 'novel tasks' and 'unrehearsed with their meanings'. (See Appendix 1).

Fifth, to the extent that different classes in a programme are formed on the basis of proficiency, proficiency scales may assist in streaming. However, in principle, it is desirable to take other factors into account besides proficiency since, in streaming, one is interested in trying to produce homogeneous learning groups and many other factors besides proficiency determine this (including, *inter alia*, age, motivation, aims, aptitude, attitudes, preferred learning strategies, and so on).

APPLICATION OF PROFICIENCY SCALES TO SCHOOL ASSESSMENT:

Finally in considering the application of proficiency scales to communicative assessment in the context of communicative-type courses in Australian Secondary Schools, one can consider what sorts of scales may be appropriate and how proficiency ratings might be translated into the sort of school grades

traditionally provided in internal and external assessment procedures in the education system.

Courses might use one of the published proficiency scales against which to assess students. This has the advantage of using instruments already extensively trialled and also has the advantage of using instruments which are widely known and which therefore facilitate comparison of student proficiency levels irrespective of the course followed and the education system experienced. There are, however, some disadvantages. Proficiency changes slowly, most proficiency scales provide no more than the 12 levels provided on the ASLPR and, like the IELTS test, often provide less, and there is little practical point in their providing more, any added differentiation being of little practical value for general purposes and increasing the risk of unreliability as the number of levels increases. Consequently, though it would seem desirable for results on a course to be interpretable in relation to one of the major proficiency scales, these scales may be considered to be too blunt for school use and to be demotivating when, for instance, the students' entry level to the course in Speaking is said to be 1 and the exit level after two years in Years 11 and 12 is said to be 1+ or 2. This may have the effect of bringing home to administrators the unintensive and therefore inefficient nature of the usual Secondary School course but the effect on learners is still demotivating. In addition, the ways in which proficiency scales name levels are not usually compatible with the ways in which education systems report results. Solutions to these problems can be found.

First, courses might, like the Queensland Secondary School language syllabuses, be accompanied by their own proficiency scales. Such scales may well be more like "graded objectives" scales related closely to the course and so reflect attainment on the course. However, they need not be unrelated to notions of general proficiency or unrelated to one of the published general proficiency scales. Indeed, courses should generally be so designed that they develop general proficiency at the same time as they specify the language content of the course reasonably precisely in terms of situations, functions, tasks, language forms and so on and this would enable quite precise and course-related scales to be provided. If the course design and scales are compatible with a general proficiency scale, then the results would be interpretable against the general scale and, for practical purposes beyond the school system, the course and the assessment can then provide information concerning the practical skills developed.

A decision may be made to use a scale which is quire course-specific and which does not lend itself to interpretation in terms of any major published proficiency scale. The language behaviour of students performing at levels A, B, C, D or E (or whatever gradings are used) may be spelt out using quite specific criteria. This is probably most easily done in courses that adopt a graded objectives approach and are quite specific in their content. Nevertheless, the problem of differentiation between the descriptors is often acute.

Another approach using a general proficiency scale such as the ASLPR is to specify the proficiency goal to be sought in the course and then to rate students at, above or below this level of attainment in whatever way is adopted in the education system in question. Thus, for instance, if the education system uses grades such as Distinction, Credit, Pass, Pass-, Fail or A, B, C, D, F, and if the goal for the average student is set as S:1+, L:1+, W:1+, R:1+, then students performing above the 1+ level would be graded at Distinction or Credit, and below at Pass-, D or F thus:

Grades	Proficiency S	L	R	W	Notes
D or A	>2	>2	>2	>2	All four macroskills >2
C or B	2	2	2	2	Two or more macroskills 2, others 1+
P or C	1+	1+	1+	1+	All four macroskills 1+
P- or D	1	1	1	1	One or more macroskill 1
F	<1	<1	<1	<1	Two or more macroskills <1

Variations below the number of macroskills at a particular level must be compensated for by an equal number above or the level below assigned.

REFERENCES

Hyltenstam, Kenneth and Manfred Pienemann. 1985. *Modelling and Assessing Second Language Acquisition*. Clevedon: Multilingual Matters.

Ingram. D. E. 1989. "Notes on the Australian Language Levels Project", notes for consideration by participants at the ALL Guidelines Evaluation Workshop, Canberra, 7 and 8 August, 1989.

Ingram, D. E. 1979, revised 1982. "Introduction to the Australian Second Language Proficiency Ratings". In Ingram 1979a rev. 1982.

Ingram, D. E. and Elaine Wylie. 1979a revised 1982. *Australian Second Language Proficiency Ratings*. Canberra: Australian Department of Immigration and Ethnic Affairs. (First edition published in 1979, revised 1982, current edition revised 1985)

Ingram, D. E. 1985. "Assessing proficiency: an overview on some aspects of testing". In Hyltenstam and Pienemann 1985: 215-276.

NAFLaSSL Project. 1989. A National Assessment Framework for Languages at Senior Secondary School Level. mimeograph.

Oller, J. W. 1976. "Evidence for a General Language Proficiency: An Expectancy Grammar". *Die Neueren Sprachen*, No. 75, 165-174.

APPENDIX 1: Speaking Scale from the Queensland Senior Italian Syllabus

SPEAKING

VERY HIGH ACHIEVEMENT

Tasks:

The student performs a variety of tasks of the complexity of an extended face-to-face conversation on everyday topics, or a talk to a specified audience on a topic of personal relevance. Registers, genres and task-types are familiar, but actual tasks are novel, with a high proportion of unrehearsed meanings. (A few minutes would be given for orientation to the situation if roleplay is involved, or for organising ideas for a monologue.)

Overall Performance:

The student performs such tasks well (though performance is far from native-like). Across the tasks, a wide repertoire of language functions and forms, including past, present and future tenses, is displayed, and there is a wide variety of sentence types, including some unrehearsed complex sentences.

Meanings Expressed:

The points essential to the task are included, and what is said is relevant and of interest to the audience. The student often goes beyond learned stereotypes to express personal perceptions, opinions and emotions, although these are substantiated or refined to only a limited degree.

Clarity and Cohesion:

A monolingual Italian listener would only occasionally need to make a special effort to understand what the student is saying or to request repetition or rewording. There is some interference from the first language in pronunciation. Links between clauses and sentences are marked by a range of simple connectors. Most inflections are correct.

Fluency:

Hesitations and circumlocutions slow down the overall rate of communication, but are not generally distracting to listeners.

Appropriateness:

Common idioms are used, and these and other expressions are appropriate in terms of formality. When the role relationships in the situation are clear-cut, the student chooses and generally sustains the appropriate second person forms.

HIGH ACHIEVEMENT

Tasks:

The student performs a variety of tasks such as those described above under VERY HIGH ACHIEVEMENT.

Overall Performance:

The student performs such tasks reasonably well. Across the tasks, a wide repertoire of language functions and forms, including past, present and future tenses, is displayed, and there is a variety of sentence types, including some unrehearsed complex sentences.

Meanings Expressed:

The points essential to the task are included. However, some of the things talked about, particularly in a monologue, are likely to be of little relevance or little interest to the audience. The student sometimes goes beyond learned stereotypes to express personal perceptions, opinions and emotions, but these are substantiated or refined to only a very limited degree.

Clarity and Cohesion:

A monolingual Italian listener would understand everything in a conversation after some repetition or rewording; in a monologue there would be some ambiguity. There is interference from the first language in pronunciation. Links between sentences and clauses are marked by simple, very familiar connectors. Most of the common inflections are correct, but there are frequent errors in less familiar forms.

Fluency:

Frequent hesitations and circumlocutions significantly slow down the overall rate of communication, and sometimes distract listeners.

Appropriateness:

Very familiar idioms are used, but these and other expressions are sometimes inappropriate in terms of formality. When the role relationships in the situation are clear-cut, the student chooses the appropriate second person forms, but does not always sustain these throughout the task, particularly with pronoun objects and possessives.

SOUND ACHIEVEMENT

Tasks:

The student performs a variety of tasks of the complexity of a simple face-to-face conversation on everyday topics, or a short talk to a specified audience on a topic of personal relevance. REgisters, genres and task-types are familiar, but actual tasks are partly novel, with some of the meanings unrehearsed. (A few minutes would be given for orientation to the situation if roleplay is involved, or for organising ideas for a monologue).

Overall Performance:

The student performs such tasks reasonably well. Across the tasks, a basic repertoire of language functions and forms is displayed. There is some variety of sentence types, though unrehearsed complex sentences are rare.

Ideas Expressed:

The points essential to the task are included. Howeve, some of the things talked about, particularly in a monologue, are likely to be of little relevance or little interest to the audience. The student expresses opinions or emotions only through learned phrases.

Clarity and Cohesion:

A monolingual Italian listener would understand all or very nearly all of the rehearsed meanings, but would miss some of the novel utterances, even after repetition or rewording. There is strong interference from the first language in pronunciation. Links between clauses and sentences are marked by only the most familiar simple connectors, e.g. e, ma, o, poi. Inflections are often wrong.

Fluency:

In unrehearsed utterances, frequent long hesitations and attempts to overcome repertoire gaps by circumlocution greatly slow down the overall rate of communication and often distract listeners.

Appropriateness:

Only the most familiar idioms are used, and these and other expressions are sometimes inappropriate in terms of formality. Even in apparently clear-cut situations the student may not choose the appropriate second person forms. If the appropriate forms have been chosen, they are not sustained throughout the task.

LIMITED ACHIEVEMENT

Tasks:

The student performs a variety of tasks such as those described above under SOUND ACHIEVEMENT.

Overall Performance:

Across the tasks, a restricted repertoire of language functions and forms is displayed. There is no ability to circumlocute, and the student may frequently resort to English words. Unrehearsed sentences consist of no more than subject, verb, and noun phrase or other simple complement. The student demonstrates very little ability to mofity or recombine learned elements effectively.

Ideas Expressed:

Only some of the points essential to tasks are included, and things talked about, particularly in a monologue, are likely to be of little relevance of little interest to the audience.

Clarity and Cohesion:

A monolingual Italian listener would understand some of the rehearsed meanings but would miss a high proportion of the novel utterances, even after a lot of repetition. Sentences are essentially unlinked.

Only the most familiar inflections are correct, and in less familiar words there may be no attempt to inflect.

Fluency:

In rehearsed utterances there are frequent, very long hesitations.

VERY LIMITED ACHIEVEMENT

Tasks:

The student performs a restricted range of tasks. These are either tasks of the complexity of those described above under SOUND ACHIEVEMENT which are very familiar, having been very recently rehearsed, or very simple tasks relating to the student's immediate circumstances, e.g. stating the number and age of members of the family.

Overall Performance:

Across the tasks, a very restricted repertoire of language functions and forms is displayed. The student demonstrates no ability to modify or recombine learned elements effectively.

Clarity:

A monolingual Italian would understand only the most rehearsed utterances.